DOCUMENT RESUME

ED 423 270 TM 029 075

AUTHOR Arter, Judy
TITLE Teaching about Performance Assessment.
PUB DATE 1998-04-00
NOTE 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Diego, CA, April 12-16, 1998).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Criteria; Educational Assessment; Elementary Secondary Education; Grading; *Instructional Effectiveness; *Performance Based Assessment; Standards; *Student Evaluation; Teaching Methods; *Test Construction; Test Use; *Testing Problems

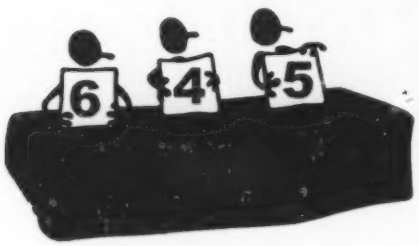
ABSTRACT

This paper describes what teachers need to know about performance assessment to do it well. First, they should know what performance assessment is and why anyone should care. One thing that has to be emphasized is that there are two parts to a performance assessment: the tasks and the criteria. Teachers tend to be better at developing rich and interesting tasks than they are at developing the criteria to describe quality performance, but it's not an assessment without both components. When to use performance assessment is another issue teachers must consider. Performance assessments are useful for assessing complex learning targets. Teachers should familiarize themselves with design options for performance assessment. Some suggestions are given for information sources about design and how to develop both tasks and criteria. In addition to considering the quality of performance assessments, teachers should consider their use as instructional methodology. The criteria can be used for teacher evaluation of students and for students' self-evaluations. Grading and reporting performance assessments is another area in which good reference material is available. It is suggested that performance assessment is a prime example of standards-based instruction, and that remembering its instructional value is very important in its use. Appendixes contain information about performance assessment criteria. (Contains 16 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 270

Teaching About Performance Assessment



PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Jerry Kirkpatrick

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Judy Arter
Northwest Regional Educational Laboratory
101 SW Main, Suite 500
Portland, OR 97204
503-275-9500
http:/www.nwrel.org arterj@nwrel.org

TM029075

Paper presented as part of the symposium: *Teaching Courses in Classroom Assessment* (Session E2, Tuesday, 2:15-3:45) at the annual meeting of the National Council on Measurement in Education, San Diego, CA April 1998

BEST COPY AVAILABLE

Teaching About Performance Assessment

Introduction

If there is *anything* definite about performance assessment, it's that we can't agree on a definition. Because of this, it's always prudent to let readers know the definition used in the current paper:

> *Performance assessment is assessment based on observation and judgment.*

One observes a performance and then judges its quality. Examples abound, everything from the driver's test and the Olympics to multitudes of daily decisions teachers make in the classroom, both summative and formative.

One can do this well or one can do this poorly. The goal of this paper is to describe what teachers need to know about performance assessment in order to do it well. And by "well" I mean prudently, efficiently, validly, and with positive consequences.

Performance assessment is not new. Teachers have been observing students and making judgments about them forever. However, there are some recent developments that highlight the importance of doing performance assessment well. For example:

1. We are asking teachers to assist students to acquire more complex skills than ever before. Just witness many of the content standards being developed by many states and professional organizations. We want students to read with comprehension, write well, be critical thinkers, be life-long learners, be collaborative workers, be able to communicate their mathematical understanding, etc. Such complex learning targets for students require complex assessments, such as performance assessments.

   Further, we are asking teachers to continuously monitor student progress toward these outcomes; it is not enough just to wait until the year-end, large-scale assessment to see how we did. This is the essence of standards-based instruction and education. So, teachers will be asked to use performance assessment on a more daily basis, and for purposes other than grading.

2. We are attempting to use performance assessment for more purposes, both inside (see #1 above) and outside the classroom (high stakes, large-scale assessment) and, therefore, there has been

   a. the attempt to make this essentially subjective form of assessment as objective as possible. Attempts along this line include standardizing tasks and criteria and careful training of those judging work, and
   b. lots of technical work on thorny issues such as sampling and generalizability.

3. Because of these changes in the uses of performance assessments and the attempts to make performance assessment more systematic, we have begun to realize that performance assessment can be used to do more than just gather numbers on students. Students can actually learn something in the process of assessing and being assessed—performance assessment materials and methods can be used to help students acquire the very skills also being assessed.

Because of these trends, there is much need to systematically teach pre- and inservice teachers about performance assessment. They do it anyway, let's help them do it well.

Now, here's my challenge to you. Do you believe that good classroom assessment will improve student achievement? Do you believe that assessing students well will make teacher's lives easier not harder? Do you believe that it is possible to use assessment as a tool for student learning as well as a tool for gathering numbers on students? Do you believe that assessment and instruction should be integrated? If we believe these things, then we have to be ready to demonstrate them to teachers. I propose that we can demonstrate all these things with performance assessment if we approach it correctly.

What Teachers Need to Know and Be Able to Do

Table 1 presents a summary of what classroom teachers need to know and be able to do with respect to performance assessment based on my experience and reading in the field.

Even though I have tried to indicate a rough order of presentation of topics, it is important to note that the knowledge and skills associated with the topics are not necessarily acquired linearly. For example, the "why we should care" part usually emerges from activities in other areas, especially how to use criteria in instruction, even though it is also emphasized right at the beginning.

As another example, consider that the notion of using performance assessment as an instructional methodology has implications for assessment design. Assessments designed for large-scale, high-stakes uses will not necessarily have the features that teachers find most useful when using performance criteria to help students learn. Therefore, adult learners almost need to have the vision of using performance assessment as a tool for learning while considering design options.

The point is that all these topics are interwoven. The trick is to decide which string would be most productive to pull on first. This is the nature of my own current learning and growth.

One more caveat before discussing some of these topics in more detail. It is impossible in a short paper to completely describe a unit on performance assessment in complete detail, including ideas on how to teach each topic. Therefore, I will try to just hit the high points and major things that seem to confuse teachers.

Table 1 Summary of Performance Assessment Knowledge and Skills		
General Topic	Specific Subtopics	Relationship to Standards
What performance assessment is and why we should care	Definitions Two mandatory parts to a performance assessment: tasks and criteria	Standards 1, 2
When to use performance assessment	Which student learning targets are best assessed with a performance assessment and which with another method Balance—performance assessment isn't always the answer Balance—ideal against practical	Standards 1
Design options	Design options for performance tasks and when to use them Design options for performance criteria and when to use them	Standard 2
How to develop tasks and criteria	Practice developing tasks and criteria	Standard 2
The nature of quality and why we should care	Quality in tasks Quality in performance criteria Consistency in scoring Sampling Avoiding possible sources of bias and distortion Building in features that result in positive consequence	Standards 1, 2, 3, 4, 7
Use as an instructional methodology	How to use performance criteria to assist students to self assess Characteristics of criteria that maximize this use Practice teaching criteria to students The role of performance criteria in standards-based education	
Grading and reporting	How to convert rubric scores to grades Ways to report on student progress besides grades	Standards 5, 6

Performance Assessment "Must Reads"

I frequently ask adult learners to read three articles: Rudner (1994), Wiggins (1992) and Stayter and Johnson (1990). The Rudner article is a balanced overview of the rationale for performance assessment, what various groups are doing with respect to performance assessment, and current issues. (Much of this is still relevant, although descriptions of activities in specific states has changed somewhat.) The message of the Wiggins paper is that quality matters. The point of the Stayter and Johnston paper is that assessment affects kids; if we want it to have a positive effect, we need to pay close attention to design issues—this affects the messages our assessments send and how we use assessment with students. These are really good papers to read at the beginning in order to set up the major themes of a unit on performance assessment.

What Performance Assessment Is and Why We Should Care

Regardless of the specific definition of performance assessment one uses, the one thing that has to be emphasized is that there are *two* parts to a performance assessment—tasks and criteria. It's not assessment if it doesn't have both. One consistently encounters "performance assessments" from supposedly reputable organizations that are just tasks. For example, a recent ASCD publication (January 1998) extolled the virtues of the following geography "authentic assessment." The teacher asked his students to research the name of their town. The students found other towns all across the US with the same name. The students wrote to each of these towns and prepared a research paper and a museum display. The "assessment" purportedly assessed research skills, geographic knowledge, and communication skills. However, there were no criteria for judging the quality of student performance on any of these skills in the context of this task. So, why is this assessment? Granted, this might be a rich, "authentic," engaging task for students, in which they might actually learn something about research skills, geography and communication. But, what? There's no way to know.

Teachers tend to be better at developing rich, interesting tasks in which to engage students than they are at developing the criteria that describe quality performance on the task. Maybe it's because we've been so process oriented in the past (rather than outcome oriented). Or maybe it's because the targets of instruction have never been clearly articulated. Regardless, good ways to illustrate the importance of criteria are:

- Give your adult learners a performance to assess; for example, show them a student giving an oral presentation. Ask them to evaluate the quality of performance on that task without giving them any criteria. Have them discuss their frustrations, ideas and solutions in small groups. Then pass out good quality performance criteria for oral presentations and ask the adult learners to again evaluate performance on the task. Have the learners discuss the difference that good criteria make.

  This points up the first need for criteria—to provide consistency between raters, and within the same rater over time and across tasks.

- Discuss the criteria used on the driver's test. (Actually call up the DMV in your state and ask them what the criteria are.) Ask your adult learners whether it would be important for students to "know" these criteria in advance, and if so, why. (There is always a resounding "YES" because criteria help students know what to practice and decrease anxiety levels.)

  Then ask your learners what it means to "know" criteria. Is it enough just to hand the criteria to the prospective driver as they begin their test? (Participants always say, "NO." For students to *know* criteria, they must be discussed and refined beginning the first day of class, practiced with feedback, be illustrated with examples of good and poor performance on each important trait or dimension of performance, modeled by the teacher, used by students to assess their own work and that of others, and used to guide the revision of work.)

- This points up the second major use for performance criteria—to help learners understand the nature of the skills they are to master and to provide a standard of comparison against which students can measure their own progress.

(Please note that such an introduction to performance assessment also begins to introduce the notion of using performance assessment both to track student progress toward important learning targets and using performance criteria as an instructional tool in the classroom.)

The major point on performance tasks that one frequently needs to emphasize is that the *task* can be any activity during the course of which the quality of performance will be observed; it doesn't necessarily have to be something that occurs separately at the end of instruction. There's a place for these summative, separate assessments, but it's also the case that much "observation and judgment" will occur during the course of regular instruction—How good is the student reading today? How good is group collaboration on this activity? This is OK. We're giving teachers permission to assess during regular instruction. The key to having daily observation be sound assessment is to have good quality criteria that teachers have internalized to the extent that they can consistently judge performances, regardless of the context. This will make daily "anecdotal records" actually mean something.

When to Use Performance Assessment

There are performance assessment zealots around. I am not one of them. For me good assessment means balanced assessment—having a clear idea of what one wants to assess and then picking the best way to assess it. My rule of thumb is: simple learning target, simple assessment; complex learning target, complex assessment. For example, knowledge and simple skills (e.g., long division) can be assessed well using multiple-choice, matching, T/F, and short answer formats. But, if we want to assess writing, mathematical problem solving, science process skills, critical thinking, oral presentations, or group collaboration skills, we'll probably need a performance assessment.

As pointed out in the introduction, teachers are being asked to assist students to acquire more complex skills and continuously monitor student progress on these skills (standards-based instruction). Therefore, teachers will need to increase their use of performance assessments.

(Many folks, however, have gotten exceedingly clever with multiple-choice questions. It's useful to show adult learners examples of attempts to assess targets like reasoning in multiple-choice format. Then show them performance assessments aimed at the same target and discuss what each format is, and is not, capable of assessing.)

Good treatments of matching methods to targets can be found in Stiggins (1997) and Marzano (1996).

Design Options

I have looked at hundreds of "performance assessments." Tasks vary in the following ways:

- A single correct answer or multiple, equally good, answers
- Group work, individual work, or a combination of both
- All written versus use of manipulatives and equipment
- Amount of choice on how to respond—written, picture, oral, etc.
- Format/length/complexity—on-demand, project, portfolio
- Amount of scaffolding—steps and processes spelled out or left up to the student
- Student choice on which task to perform

Performance criteria vary in the following ways:

- Task-specific (criteria spell out what responses should look like on each task separately) versus general (the same criteria are used across similar tasks)
- Holistic (one score for the whole performance) versus analytical trait (several scores for a single performance based on the important dimensions of the performance; for example in writing one might assess the quality of the ideas, organization, voice, word choice, sentence fluency, and conventions)
- Number of score points (I've seen from three to nine)

The question to be addressed is: Why do developers make these choices? What are the advantages and disadvantages of each? Choices are usually made by balancing ideal against practical and throwing in technical considerations for good measure. What is usually left out when considering design options are potential impacts on teachers and students, including the idea that, if designed properly, performance assessment materials, especially criteria, can be used instructionally.

As one example, many large-scale performance assessment developers use task-specific, holistic criteria because (a) it is faster to train raters; (b) it is faster to score; and (c) it is believed that such criteria will result in high agreement rates among raters. More useful in the classroom,

however, are generalized, analytical trait criteria because they assist teachers and students to articulate the features of sold work that can be generalized from task to task. The value of general criteria for instruction has been noted in a recent study by OERI (1997).

A good discussion of design options with lots of examples and some discussion on why and when to use various options can be found in Regional Educational Labs (1998).

How to Develop Performance Tasks and Criteria

Two of my favorite sources here come from Jay McTighe's work with the Maryland Assessment Consortium (McTighe, 1996) and some of the training activities in the resource notebook developed by the Regional Educational Laboratories (1998).

The Nature of Quality and Why We Should Care

Actually, we've been talking about quality all along. Picking the correct assessment method and purposefully designing tasks and criteria are all part of quality performance assessment. A summary of these and other quality considerations for performance assessments (summarized from our own experience as well as what a lot of other authors have said) is included in *Appendix A* (from the Regional Educational Labs, 1998).

Many good papers on the current state of the art are available. One of my favorites is Herman (1996). These studies show that we have a long way to go make our performance assessments work well for high-stakes uses.

Coming up.with a list of quality features of performance assessment is the easy part. Teaching adult learners the nature of quality and why we should care is the tricky part. In my experience, what doesn't work is to show learners the list of quality considerations and lecture one's way from beginning to end. What seems to work better is to engage adult learners in a good performance assessment (try the six-trait model for assessing writing; Spandel and Stiggins, 1997) and a bad performance assessment (try **Activity 1.5**—*Clapping Hands* from the Regional Educational Labs, 1998) and let them tell you what the differences are. Teachers more readily understand sampling or bias or unclear tasks when they experience the effect these things have on themselves.

Activity 1.5 (mentioned above) is designed to let learners construct for themselves the meaning of quality and why they should care by engaging them in a bad performance assessment and having them articulate the effect that such an experience has on them. From this activity participants generate a simplified version of the criteria in *Appendix A*.

Then, it's always important to practice applying the criteria to samples. So, find good and poor performance assessments and let learners practice critiquing them as a group. Ask your adult learners what advice they would give the author to make the assessment better, and then have them work in groups to improve the assessment using their own suggestions. When your adult

learners have had experience with this, have them develop their own tasks and criteria and work in peer review groups to improve their own efforts. Maybe even have learners keep a portfolio on their efforts and reflect on how their assessments have improved with time.

The interesting part of this process is that this is exactly the same way to teach *any* criteria to *any* group of students; so, in essence, you're modeling for prospective or active teachers what you want them to be doing with their own classes in the future.

At this point, I need to mention again that statements of quality must take use into account. For example, as described above, the features most desirable in criteria used for a large-scale, high stakes assessment might not be those most desirable in a classroom assessment that will also be used instructionally. For classroom use, general, analytical trait criteria with levels clearly describe in terms that students can understand, are necessary.

Regional Educational Laboratories (1998) also contains other packaged instructional activities designed to help adults learn about the meaning of quality as applied to performance assessments, and describes criteria for criteria for both large-scale and instructional uses.

Use as an Instructional Methodology

I've alluded to this topic several times before in this paper. There's lots to say here, but I won't. My goal here is just to give the flavor of the possibilities and how to help teachers have these possibilities become realities.

As mentioned before, good quality, *general* performance criteria can be a powerful instructional tool (OERI, 1998; Arter, et. al, 1994). The idea is that students have a much better chance of hitting a target (and teachers have a much better chance of teaching to the target) if it is well defined. What is critical thinking, and how will we know when students are competent at it? What is life-long learning, and how will we know when students are capable of doing it? What do competent writing or oral presentations or group collaboration look like?

Teachers are being asked to assist students to acquire such complex skills. Many of these are ill defined and fuzzy to teachers. I am proposing that the attempt to develop and use performance criteria that define these constructs will enormously assist teachers to know what to teach and students what to learn. I'm sure that many of you have had this experience, usually with respect to writing, but maybe for other skills as well, such as math problem solving or communicating mathematical understanding. Two papers that articulate well the power of using general performance criteria to help students learn are Spandel (1997) and Arter (1996).

The quotes in *Appendix B* also illustrate this point well.

There are two steps in making performance criteria work as tools for learning in the classroom: having a notion in our own minds what the criteria are, and then teaching them to students.

Step 1: Having a notion in our minds of what the criteria are. There are three ways to help adult learners clarify criteria.

First, if the person is an expert in the field for which criteria are being developed, he or she can sometimes just sit down and write the criteria out. This process can be illustrated with many common everyday situations. For example, most adults have fun developing criteria for effective whining, room cleaning, or restaurants.

In education, most of us are familiar enough with oral presentations that we stand a chance of articulating and writing out decent criteria. And measurement experts usually can write out the criteria for a good quality performance assessment.

These "off the top of my head" type criteria, however, need to be subjected to a reality test. After the criteria are written down, it is always a good idea to gather samples of the performance under consideration and try to rate them using one's criteria. This process helps one notice important aspects of performance one forgot, borderline cases that need to be clarified through refined statements in the criteria, the need to more clearly specify levels of performance with indicators of quality, etc. (For example, notice that our criteria for a good quality performance assessment in *Appendix A* don't have levels defined; and, will everyone agree on what each of the items listed means? *Appendix A* still needs some work.)

The *second* way to develop general performance criteria is to just start with the student work; and, this is actually where most teachers begin. Get sets of student work that illustrate various levels of quality on the skill in question—writing, communicating mathematical understanding, critical thinking, etc. Ask adult learners to sort them into three stacks—strong, medium and weak performances. Then have them begin to describe the differences between the stacks. This method is illustrated in detail in **Activity 2.1**—*Sorting Student Work* (Regional Educational Laboratories, 1998). Seven samples of student work from various grade levels and content areas are also included (see Appendix B of the previous reference)

If your adult learners have trouble sorting work, it means that they don't have enough of an idea of the construct in question to have formed even intuitive criteria. This means the *third* approach kicks in; they need to hit the literature in the content area in question. This frequently occurs, for example, when teachers are attempting to develop criteria for critical thinking. This also may be the case for preservice teachers in many content areas, constructs and skills. Defining the construct and collecting relevant samples of student work to sort probably provides a way to good interface with the content area courses preservice teachers are taking.

Even though actually developing criteria is a good exercise, it is fortunate that we don't have to start from scratch all the time. There are many good sources of criteria and rubrics. One of my favorites is Perlman (1994). The 70 or so rubrics in this collection provide a good opportunity for adult learners to practice distinguishing good criteria from weak ones. (And for practicing distinguishing criteria that might work for large-scale assessment from those that might be most useful in the classroom.)

Step 2: Teach students to self assess using criteria. The second step in making performance criteria work as instructional tools in the classroom is to teach them to students. My colleagues have developed seven strategies for teaching criteria to students (Spandel and Stiggins, 1997; Spandel and Culham, 1998). These were developed in the context of writing, but many of the strategies transfer easily to other skills. The seven strategies are:

1. Teach students the vocabulary they need to think and speak like writers. Help students understand the nature of quality through engaging them in the same sort of sorting and descriptive activities described above. In other words, engage students in developing criteria for quality.

2. Read, discuss and score anonymous samples of student work. Once criteria are in place, students need to practice using them to notice what is strong and weak in work. They need to not only judge the overall quality of work, but also articulate the reasons for their judgment of quality. We usually use highly descriptive criteria written in student language. To justify their scores, we ask students to find the words in the criteria that describe the paper under consideration.

3. Practice focused revision. In addition to being able to notice what is strong or weak in work, students need to know what to do fix something that is weak. One procedure is to have them give advice (as a group) to the author, and then work in groups to improve the sample performance using the advice given. For example, when working on the trait of ideas in writing, students might notice that the paper is weak because it is unfocused, doesn't stick to the point, emphasizes irrelevant details, or tells rather than shows (all descriptors in the criteria we use). We would then ask students to revise the writing using their own advice.

4. Use samples of work to illustrate the criteria. For writing, this can be done with various published items—stories, picture books, manuals, instructions, etc.

5. Have students help you revise your work using the criteria for quality.

6. Allow students many opportunities to articulate their emerging notion of quality—write letters to authors, describe their progress to parents, teach younger children, etc.

7. Teach mini-lessons. For writing, this means organizing regular instructional lessons by trait: ideas, organization, voice, word choice, sentence fluency and conventions. If the criteria really do describe what we mean by quality, why not teach directly to them? Teachers already teach the traits of good writing. What is many times lacking is the conscious link between what is being taught and the standards being developed.

Grading and Reporting

This topic will be covered in the next paper by Susan Brookhart and so will not be discussed at length here. However, there are many helpful papers and training activities available to assist adult learners be good graders and reporters of student achievement. One good source for instructional activities is Regional Educational Laboratories (1998). The ASCD Yearbook on Grading and Reporting (1996) is also a great source of current thought on the topic of grading and reporting.

Performance Assessment and Standards-based Instruction

I have begun asking participants in my workshops (teachers, administrators, and others) to state the most salient characteristics of "standards-based instruction" as they currently understand it. Admittedly an unscientific sample, here is the essence of what they say. Standards-based instruction/education is characterized by:

- There are clearly stated long-term learning targets for students (content standards). These articulate what we want students to know and be able to do when students leave K-12 education.
- There are benchmarks along the way so that we know if we are on track getting students to the ultimate level of competency.
- These learning targets are connected to the real world.
- Instruction and assessment are aimed at these important targets and aligned across grade levels to reduce duplication and make it clear how the skills and understandings developed one year will be built upon the next.
- We use standards-referenced descriptions of student learning, rather than norm-referenced or self-referenced—we define the nature of quality and match student achievement to it in order to judge student achievement.
- Everybody—students, teachers, parents, community members—is aware of the nature of excellence and what it will take to succeed. Students can plot where they are.

I submit that performance assessment is a prime example of such standards-based instruction.

- Performance criteria help define the standards—they are the final definition.
- The process of developing performance criteria is more than an exercise in assessment; it helps instruction.
- Performance criteria make standards clear to students.
- Teaching criteria to students improves the very skills we're also assessing—integrating assessment and instruction.

References

Arter, J. (1996). Using assessment as a tool for learning (Paper IV-10). In R. Blum and J. Arter (Eds.), *Student performance assessment in an era of restructuring.* Alexandria, VA: Association for Supervision & Curriculum Development.

Arter, J., Spandel, V., Culham, R., & Pollard, J. (1994). The impact of training students to be self-assessors of writing. Paper presented at AERA, New Orleans, 1994.

ASCD Year Book (1996). *Communicating student learning* (T. Guskey, Ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Herman, J. (1996). Technical quality matters (Paper I-7). In R. Blum and J. Arter (Eds.), *Student performance assessment in an era of restructuring.* Alexandria, VA: Association for Supervision & Curriculum Development.

Marzano, R.J. (1996). Understanding the complexities of setting performance standards (Paper I-6). In R. Blum and J. Arter (Eds.), *Student performance assessment in an era of restructuring.* Alexandria, VA: Association for Supervision & Curriculum Development.

McTighe, J. (1996). Performance-based assessment in the classroom: A planning framework (Paper I-5). In R. Blum and J. Arter (Eds.), *Student performance assessment in an era of restructuring.* Alexandria, VA: Association for Supervision & Curriculum Development

OERI (1997). Studies of education reform: Assessment of student performance. (Authors: N. Khattri, A.L. Reeve, R.J. Adamson). Washington DC: Office of Educational Research and Improvement.

Perlman, C. (1994). The CPS performance assessment idea book. Chicago Public Schools, 1819 W. Pershing Rd., Chicago, IL 60609.

Regional Educational Laboratories (1998). *Improving classroom assessment: A toolkit for professional developers,* 2nd edition. Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, OR 97204.

Rudner, L., & Boston, C. (1994). Performance assessment. *ERIC Review,* Vol. 3(1), pp.2-12.

Spandel, V. (1996). Criteria: The power behind revision. *Writing Teacher,* January, pp. 9-25.

Spandel, V. and Culham, R. Writing workshop materials. Personal communication, Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, OR 97204.

Spandel, V. & Stiggins, R.J. (1997). *Creating writers: Linking writing assessment and instruction.* New York: Longman.

Stayter, F., & Johnston, P. (1990). Evaluating the teaching and learning of literacy. T. Shanahan (Ed.), *Reading and writing together: New perspectives for the classroom,* Norwood, MA: Christopher-Gordon.

Stiggins, R.J. (1997). *Student-centered classroom assessment,* 2nd edition. Columbus, OH: Merrill.

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership,* May, pp. 26-33.

Appendix A: Criteria for Performance Assessments[1]

	Yes	Somewhat	No
1. Content/Skill Coverage and Correct Method	3	2	1

The assessment:

- Clearly states skills and content to be covered
- Correctly uses alternative assessment to measure these skills and content
- Avoids irrelevant and/or unimportant content
- Deals with enduring themes or significant knowledge
- Matches statements of coverage to task content and performance criteria

2. Performance Criteria:	3	2	1

- Include everything of importance and omit irrelevant features of performance
- State criteria clearly and provide samples of student work to illustrate them
- Are stated generally, especially if the intent is use as an instructional tool
- Are analytical trait, especially if the intent is use as an instructional tool

3. Performance Tasks:	3	2	1

General:

- Elicit the desired performances or work;
- Recreate an "authentic" context for performance
- Exemplify good instruction
- Are reviewed by others (students, peers, experts)

Sampling/Representativeness/Generalizability:

- Cover the content or skill area well; results can be generalized
- Sample performance in a way that is representative of what a student can do

Bias and Distortion:

- Avoid factors that might get in the way of student ability to demonstrate what they know and can do

---

[1] From Regional Educational Laboratories (1998), Chapter 3, pp. 40-41.

	Yes	Somewhat	No
4. Fairness and Rater Bias	3	2	1

Performance Tasks:

- Have content and context that are equally familiar, acceptable, and appropriate for students in all group
- Tap knowledge and skills all students have had adequate time to acquire in class
- Are as free as possible of cultural, ethnic, or gender stereotypes
- Are as free as possible of language barriers

Performance Criteria and Rater Training:

- Ensure that irrelevant features of performance do not influence how other, supposedly independent, features are judged
- Ensure that knowledge of the type of student does not influence judgments about performance quality
- Ensure that knowledge of individual students does not affect judgments about performance quality

5. Consequences	3	2	1

The assessment:

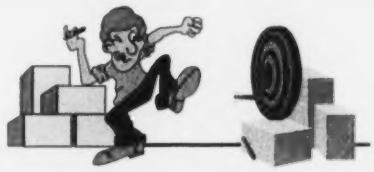
- Communicates appropriate messages
- Results in acceptable effects on students, teachers, and others
- Is worth the instructional time devoted to it; students learn something from doing the assessment and/or using the performance criteria
- Provides information relevant to the decisions being made
- Is perceived by students and teachers as valid

6. Cost and Efficiency	3	2	1

The assessment:

- Is cost efficient—the results are worth the investment
- Is practical/"do-able"

Appendix B: Performance Criteria Quotes—Why Performance Criteria are Important



[Students have] been conditioned to believe that great papers "just happen," that they are a guessing game, and that one finds out what to do after it's too late.

Margie Krest, *Adapting the portfolio to meet student needs,* **English Journal,** February 1990

I want [students] to see evaluation in its best sense—a source to inform teaching and learning. To that end we develop a vocabulary for commenting on the admirable and problematic aspects of writing....The more we examine samples, the richer and more helpful this language of evaluation becomes.

Melinda Erickson, *Developing student confidence to evaluate writing,* **Quarterly Journal of the National Writing Project,** 1992, pp. 7-9

Winning points may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game...however often it seems to be about scoring more wins than losses.

Catherine Lucas, *Introduction: Writing portfolios—changes and challenges.* In K. Yancey, **Portfolios in the writing classroom**, NCTM, 1992, pp. 1-11.

Scales, criteria, and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them. These treatments are two times more effective than free writing techniques.

George Hillocks, Jr., *Research on written composition*, NCTE, 1986.

Ultimately, we want students to grow to be independent. For them to do that, they have to have a sense of what the criteria are that make them successful. For a long time, the criteria have been a mystery to students.

Robert Tierney, et. al, **Portfolio assessment in the reading-writing classroom**, NCTE, 1991.

Teachers [frequently] ask the wrong question first..."What do we do?"—putting the focus immediately on designing tasks—when they need to ask, "What do we want kids to know and be able to do? How well? What does quality look like?" [We] need to ask these questions very clearly first.

Mike Hibbard, **Education Update**, Vol. 38(4), p. 5, ASCD, June 1996.

While the purpose of student selection [in portfolios] is to engender and support a reflective and self-evaluative capacity, the developers recognized that this is possible only if students have deep understandings about the nature of quality in their work and are able to make judgments that accurately reflect a valid assessment of that quality.

Paul LeMahieu, Drew Gitomer, and JoAnne Eresh, *Portfolios in Large-Scale Assessment: Difficult But Not Impossible*
Educational Measurement: Issues and Practice
Fall, 1995, pp. 11-28.

In my work with teachers, it is the skills of "rubric writing" which is most elusive. Perhaps it's because we're used to assigning single grades for complex assignments, knowing what an "A" looks like in our heads, but rarely "putting it to paper" so that our students can see it as well. Perhaps the difficulty in writing scoring criteria also lies in the challenge of describing just what it really looks like to perform well, or better yet, to perform at a variety of levels of competency.

Nevertheless, it is the use of rubrics as an indispensable part of the instructional process which completes the vital link between assessment and instruction. Until we invest the time discerning for ourselves what excellence in writing, or speaking, or dancing, or singing, or whatever looks like, we are unable to fully "teach" our students to achieve at these levels.

Scott Mendel, *Creating portraits of performance*, Peakview Elementary School, 19451 East Progress Circle, Aurora, CO 90015